

IPEDS Analytics: Delta Cost Project Database 1987-2010

Data File Documentation

AUGUST 2012

NCES 2012-823
U.S. DEPARTMENT OF EDUCATION



This page intentionally left blank.

IPEDS Analytics: Delta Cost Project Database 1987-2010

Data File Documentation

AUGUST 2012

Colleen Lenihan
National Center for Education Statistics

NCES 2012-823
U.S. DEPARTMENT OF EDUCATION



U.S. Department of Education

Arne Duncan
Secretary

Institute of Education Sciences

John Q. Easton
Director

National Center for Education Statistics

Jack Buckley
Commissioner

The National Center for Education Statistics (NCES) is the primary federal entity for collecting, analyzing, and reporting data related to education in the United States and other nations. It fulfills a congressional mandate to collect, collate, analyze, and report full and complete statistics on the condition of education in the United States; conduct and publish reports and specialized analyses of the meaning and significance of such statistics; assist state and local education agencies in improving their statistical systems; and review and report on education activities in foreign countries.

NCES activities are designed to address high-priority education data needs; provide consistent, reliable, complete, and accurate indicators of education status and trends; and report timely, useful, and high-quality data to the U.S. Department of Education, the Congress, the states, other education policymakers, practitioners, data users, and the general public. Unless specifically noted, all information contained herein is in the public domain.

We strive to make our products available in a variety of formats and in language that is appropriate to a variety of audiences. You, as our customer, are the best judge of our success in communicating information effectively. If you have any comments or suggestions about this or any other NCES product or report, we would like to hear from you. Please direct your comments to

NCES, IES, U.S. Department of Education
1990 K Street NW
Washington, DC 20006-5651

August 2012

The NCES Home Page address is <http://nces.ed.gov>.
The NCES Publications and Products address is <http://nces.ed.gov/pubsearch>.

This publication is only available online. To download, view, and print the report as a PDF file, go to the NCES Publications and Products address shown above.

Suggested Citation

Lenihan, C. (2012). *IPEDS Analytics: Delta Cost Project Database 1987-2010* (NCES 2012-823). U.S. Department of Education. Washington, DC: National Center for Education Statistics. Retrieved [date] from <http://nces.ed.gov/pubsearch>.

Content Contact

Colleen Lenihan
(202) 502-7481
Colleen.Lenihan@ed.gov

Overview

The IPEDS Analytics: Delta Cost Project Database was created to make data from the Integrated Postsecondary Education Data System (IPEDS) more readily usable for longitudinal analyses. Currently spanning the period from 1987 through 2010, it has a total of 202,800 observations on 932 variables derived from the institutional characteristics, finance, enrollment, completions, graduation rates, student financial aid, and human resources IPEDS survey components and a limited number of outside sources.

The maintenance and hosting of the IPEDS Analytics: Delta Cost Project Database was taken over by the National Center for Education Statistics (NCES) in 2012. The database was originally created by the Delta Cost Project (an independent, nonprofit organization) in 2007. For a detailed history of the development of the database under the Delta Cost Project, which covers the 1987-2009 database, please refer to its location on the NCES website, http://nces.ed.gov/ipeds/deltacostproject/download/DCP_History_Documentation.pdf .

The database has been posted online in two parts for easier downloading; the first part contains the file for the 1987-1999 academic years and the second for the 2000-2010 academic years. These files are intended to be merged together to create the full 1987-2010 database.

Design

The IPEDS Analytics: Delta Cost Project Database was created to make IPEDS data more readily usable for longitudinal analyses. The database has been organized to have one observation per institution for each year. The database includes data for every institution that has reported institutional characteristics data to IPEDS for the fall of the academic year. These data have been harmonized in order to mitigate changes in financial reporting standards over time by employing industry-accepted manipulations of the data. When possible, missing data have been replaced via imputation. The database has been organized to further ease longitudinal analyses by creating consistent institutional groupings and matched sets to account for changes to the IPEDS universe of institutions over the time period. Additionally, variables to adjust the financial information to constant dollars have been included for the Consumer Price Index-Urban Consumers (CPI-U), the Higher Education Price Index (HEPI), and the Higher Education Cost Adjustment (HECA).

Institutional Groupings

NCES allows certain institutions (“parent institutions”) to report data for branch campuses or other affiliated institutions (“child institutions”) for various IPEDS surveys. Parent institutions may have one or more child institutions and these parent/child relationships may differ over time and/or by survey. The need for this combined reporting often depends on the type of survey—child institutions may report their own data on some surveys (e.g., enrollment or completions), while the parent institution reports their combined data on other surveys (e.g., finance). These reporting relationships can also change when affiliated institutions are opened or closed, so the parent/child reporting structures may change over time and/or cease to exist.

Institutions that reported data together due to having a parent/child reporting relationship on any of the IPEDS surveys for any year between 1987 and 2010 have been grouped together for all years in order to maintain the consistency of the data for the entire time period. This means that all of the data for these parent/child institutions has been combined to make one observation per year for the set of institutions. The exact number of groupings in the database fluctuates from year to year; for the 2010 academic year, there were 567 institutional groupings in the dataset. Of these institutional groupings present in the 2010 academic year, 168 are public, 121 are private nonprofit, and 278 are private for-profit.

Longitudinal Institution Panels

In order to ensure that trends in the data are not being affected by institutions coming into or leaving the dataset of analysis, the database includes variables to identify panels of institutions that report data consistently over specified time periods. These institutional panels, referred to as “matched sets,” have been created for U.S. public and private nonprofit 4-year and 2-year institutions that are classified as Associate’s, Baccalaureate, Master’s, and Research institutions according to the Carnegie 2005 Classifications. In order to be included in the matched set, the institution must have data on fall full-time equivalent (FTE) student enrollment, instructional

expenditures, and student completions for every year of the time period. There are three different matched sets to cover different time periods: 1987-2010, 2000-2010, and 2005-2010. Institutions that have extreme outlier data in the time period or that have changed sector or Carnegie Classification have been removed from the pertinent matched set.

The table below shows the institution counts for the three matched set panels for institutions in the seven major Carnegie/sector classifications.

Carnegie Classification 2005 by Sector	2005-2010 6-year matched set (matched_n_05_10_6)	2000-2010 11-year matched set (matched_n_00_10_11)	1987-2010 24-year matched set (matched_n_87_10_24)
Public Research	152	152	151
Public Master's	230	230	228
Public Bachelor's	89	86	83
Public Associate's	833	819	703
Private Nonprofit Research	100	99	97
Private Nonprofit Master's	313	311	304
Private Nonprofit Bachelor's	470	466	440

Data Harmonization

The Delta Cost Project has harmonized the IPEDS finance data to provide comparable revenue and expenditure data over time and across different financial reporting standards, to the extent possible. These adjustments ensure reasonable consistency in the patterns over time and allow broad comparisons between public and private institutions. In the standard IPEDS data, many of the finance variables are not consistent over time due to changes stemming from the conversion from the Common Form reporting format to separate Governmental Accounting Standards Board (GASB) and Financial Accounting Standards Board (FASB) reporting formats. The variables provided in the IPEDS Analytics: Delta Cost Project Database include the original data reported in IPEDS as well as the adjusted versions that have been used by the Delta Cost Project in their trend analyses.

For revenues, the most notable adjustments are to net tuition, federal grants and contracts, and auxiliary enterprise revenues. These adjustments have been made to account for the inconsistencies caused by reporting revenue amounts net of "applied discounts and allowances" under FASB, and later, GASB reporting standards. Over the entire 1987-2010 period, the net tuition amount in the Delta Database has been standardized to be gross tuition revenue net of only institutional grant aid. Federal grant revenues have been adjusted to be net of Pell Grants (where applicable), as these are captured in the net tuition revenue amounts. Sales and service of auxiliary enterprise revenues are provided in gross amounts only.

For expenses, adjustments have been made to the functional expenditure categories to account for changes in the reporting of Operations and Maintenance (O&M) and Interest across different reporting standards. Following Common Form and GASB reporting formats, O&M and interest

were separate expenditure categories; under the FASB and New Aligned form reporting formats these amounts had to be embedded in the other functional expenditure categories. The main expense variables in the database have been calculated for these amounts to be consistent over time by subtracting O&M and interest from the functional expenditure categories and then summing those O&M and interest amounts separately to create variables representing total amounts.

In addition to adjusting the data to be comparable across accounting standards, the data have also been organized to translate accounting information into more commonly understood data elements that reflect practical information for institutions and policy audiences. Revenue variables have been derived to show the amount of money coming from students, public sources, and private sources that are generally at the institution's discretion to determine how these funds are spent as opposed to those revenues that are restricted to certain purposes (such as hospitals and independent operations). Revenue variables have also been put in the context of expenditures to show the portion of educational expenses that come from students against those expenses that are subsidized by the institution.

Expenditure variables have been derived to present the functional expenditure variables in the broader context of different institutional purposes. Instruction, student services, and the associated share of overhead costs are grouped into education and related expenses; research and the associated share of overhead costs are grouped into research and related costs; and public service and the associated share of overhead costs are grouped into public service and related costs. These three categories along with net scholarships and fellowships combine to be education and general spending. The expenditures that are largely self-supporting, including independent operations, auxiliary enterprises, and hospitals are aggregated into a separate category. Variables have also been derived to put expenditures into the context of completions to show an estimate of what an institution spends for each degree or completion in a given year.

Imputations

The Delta Cost Project IPEDS Database involves two different imputation procedures. The first imputation procedure utilized a conservative methodology to fill in gaps for missing data for the general dataset. The second imputation procedure was done to account for changes in reporting standards over time for institutions following FASB accounting standards.

To develop a more robust dataset, regression imputation procedures have been employed as needed for all variables. Delta adopted a relatively conservative method to impute data for an institution any time that there was a 1-year gap between two data values (e.g., missing 2003 data for a series would be imputed for if there were data for 2002 and 2004). If the gap between values was 2 years or more, the gap was not filled in. Furthermore, values were not imputed when there were missing data at the beginning or end of the data series for an institution. There are imputation flags in the database to denote any instance where a value has been imputed.

A second imputation procedure was developed to improve the comparability between Common Form, FASB, and GASB expenditure data. In this methodology, data were imputed for FASB-reporting institutions when institutionally reported data were unavailable from 1997 to 2003. Interest and O&M expense data were not reported for any FASB institution between 1997 and 2001, therefore each was separately imputed. This imputation process was also employed for institutions that did not report interest or O&M data (or reported partial data) for 2002 and 2003.

The specific methodology for imputing the missing interest and O&M data from 1997 to 2003 used data that was reported from 2002 to 2008. First, the reported interest and O&M in each functional expense category were computed separately as a share of total expenditures. Then, for each institution, an institutional median share was also determined for interest and O&M for each expense category across the 2002-2008 period; the institutional median was used in years when there was no reported share. For those institutions with no reported data for a particular expense category over the 2002-2008 period, a “peer group median share” was constructed using the median share from a set of institutions with the same Carnegie Classification and similar FTE and core expenditures (instruction, student services, academic support, and institutional support). The shares for interest and O&M (institutional shares, institutional median shares, or peer group median shares) were then applied to the total expenditures for all years, 1997-2003; imputed values were assigned where interest and O&M data were missing. The sum of the interest and O&M data for each functional category were then scaled to ensure they summed to the control totals for interest and O&M.

For a more detailed history of the development of the database, including data harmonization, groupings, imputations, and other processing issues from the 1987-2009 database, please refer to http://nces.ed.gov/ipeds/deltacostproject/download/DCP_History_Documentation.pdf.

Cautions to Users

NCES assumed control of the Delta Cost Project IPEDS Database 1987-2009 with the understanding that NCES would: (a) provide annual updates to the database to bring in new data as it becomes available, (b) update institutional groupings as necessary, and (c) provide imputations for data missing from the prior year where possible.

Upon receipt of the database, NCES reviewed its contents for compliance with NCES Statistical Standards. In so doing, a limited number of inconsistencies were noted. These include: (a) percentage or share values that do not sum to 100 percent, (b) imputed values that are outside of the expected range, and (c) negative values where a negative amount is not feasible.

The majority of these inconsistencies appear related to imputation, specifically affecting variables where both total amounts and component parts are included in the database. Delta Cost Project imputation methodology did not consistently force the reconciliation of imputed component amounts to match reported totals, or vice versa. For example, if a component amount, such as salary expenses for academic support, has been imputed, then it is possible for this amount to be greater than the total amount reported for academic support expenses as

a whole. While it is rare for this mismatch to happen, it is possible using the Delta Cost Project imputation methodology and can result in unreasonable values for derived variables. NCES followed the Delta Cost Project methodology for the 2010 database update, including the imputations for data missing in 2009. In future updates of the database, the imputation methodology will be revised to reconcile the imputed amounts.

File Updates for the IPEDS Analytics: Delta Cost Project Database 1987-2010

This section contains a summary of the changes incorporated into the IPEDS Analytics: Delta Cost Project Database 1987-2010. The changes described include those made since the 1987-2009 file was released (on August 23, 2011) in addition to importing the 2009-10 IPEDS data into the database.

Changes to the 1987-2010 data file

1. New Variables

Variable	Label	Notes
total_enrollment_multi_tot	Total enrollment (Multi)	NCES started collecting information on enrollment of students that identify as being more than one race. This information was optional starting in the fall of academic year 2009 and will be mandatory for academic year 2011.

2. Revised Variables

- *Inflation Variables*

Variables	Revision
CPI_Scalar_2010	
HEPI_Scalar_2010	
HECA_Scalar_2010	The scalar variables were recalculated to inflate dollars to 2010 constant dollar amounts rather than 2009 constant dollar amounts and were renamed to reflect this change.

- *Matched Set Variables*

Variables	Revision
matched_n_87_10_24	
matched_n_00_10_11	
matched_n_05_10_6	The matched set variables were advanced a year to reflect an additional year of data. The number of institutions in the matched set will vary depending on whether Carnegie_sector_2000 or Carnegie_sector_2005 is used for analysis, as some institutions changed categories in the Carnegie 2000 and Carnegie 2005 classifications. The matched set variables only includes institutions in the United States (institutions located in territories are not included) that have consistently reported data on instructional spending, fall full-time equivalent student enrollment, and completions. Some institutions with complete data were removed from the matched set because they contained extreme outliers.

Revised institutional groupings

Any time an institution is a “parent institution” and has a new “full child” institution included in its data, these institutions are grouped together in the database. As long as the new “full child” institution has never reported its own separate data to IPEDS, the inclusion of the institution’s data with the parent institution’s data does not change the information that has been previously included in the data file.

Occasionally, institutions that have previously reported separate data merge together with the result that their data need to be grouped for the entire span of the database, which does change the data for these institutions from the 1987-2009 data file. The table below lists the institutions that have merged and now have revised grouped data in the database.

Institutional grouping	GroupID	Institutions included	UnitIDs
Santa Clara University	2900	Santa Clara University Jesuit School of Theology at Santa Clara	122931 116624
Middlebury College	2901	Middlebury College Monterey Institute of International Studies	230959 119058
University of Connecticut	2050	University of Connecticut System (previously grouped) University of Connecticut Medical and Dental School	129020 243762